

# Parameter Choice by Fast Balancing

**Frank Bauer**

E-mail: frank.bauer.de@gmail.com

**Abstract.** Choosing the regularization parameter for inverse problems is of major importance for the performance of the regularization method.

We will introduce a fast version of the Lepskij balancing principle and show that it is a valid parameter choice method for Tikhonov regularization both in a deterministic and a stochastic noise regime as long as minor conditions on the solution are fulfilled.

AMS classification scheme numbers: 47A52,65J22,60G99,62H12

## 1. Introduction

In the following we will consider linear inverse problems [EHN96, Hof86] given as an operator equation

$$Ax = y, \quad (1)$$

where  $A : \mathcal{X} \rightarrow \mathcal{Y}$  is a linear, continuous, compact operator acting between separable real infinite dimensional Hilbert spaces  $\mathcal{X}, \mathcal{Y}$ . Without loss of generality, we assume that  $A$  has a trivial null-space  $N(A) = \{0\}$ . Since  $A$  is compact and  $\mathcal{X}$  is infinite dimensional,  $A$  does not have a continuous inverse, which makes (1) ill-posed.

For some definitions, but not for the methods themselves, we will need the singular value decomposition of  $A$ . There exist orthonormal bases  $(u_k)_{k \in \mathbb{N}}$  of  $\mathcal{X}$  and  $(v_k)_{k \in \mathbb{N}}$  of  $\mathcal{Y}$  and a sequence of decreasing singular values  $(\sigma_k)_{k \in \mathbb{N}}$  such that

$$Ax = \sum_{k=1}^{\infty} \sigma_k \langle x, u_k \rangle v_k. \quad (2)$$

Moreover, we assume that the data  $y$  are noisy, the noise model for  $\xi$  will be specified later.

$$y^\delta = Ax + \xi, \quad \xi \text{ noise.} \quad (3)$$

In order to counter the ill-posedness, we need to regularize; in this article we will consider only Tikhonov regularization:

$$x_n^\delta = A_n^{-1} y^\delta := (A^* A + q_0 q^n)^{-1} A y^\delta \quad (4)$$

The level  $n$  will now be called regularization parameter;  $q_0 > 0$  and  $0 < q < 1$  are constants which are discussed later. The noise-free regularized solution is defined as

$$x_n = A_n^{-1} y = (A^* A + q_0 q^n)^{-1} A y \quad (5)$$

The correct choice of the regularization parameter is of major importance for the performance of the method. In recent times, a number of articles [GP00, MP03, BP05, BH05, MP06, HPR07, BHM09] have considered the Lepskij Balancing principle [Lep90] for choosing this parameter in various situations.

One of the major disadvantages of this parameter choice method in comparison to, for instance, the Morozov Discrepancy principle (cf. e.g., [EHN96]) is that one needs to compute all regularized solutions up to a maximal regularization parameter. On the other hand, this buys stability even in the face of stochastic noise.

We will show that a simplification of the Lepskij balancing principle (now called fast balancing) will yield a valid parameter choice method which performs at least as well as the original. This idea has already been presented in a different form in [RH08], however in a purely deterministic setting with a focus on convergence results. In contrast our goal is to provide feasible error bounds in realistic situations without requiring the noise level to be near to or converge to 0. There are methods (see e.g. [HPR08]) which reach by a combination of solutions a better solution, however as the impact to practice has

been rather small still, we will keep our focus to choose the best solution in a set of given ones.

The outline of this paper is as follows. First (section 2) we will specify the conditions on the solution  $x$  and describe two different scenarios of noise, namely a stochastic and a deterministic one. In section 3 we will introduce the Lepskij balancing principle and the fast balancing principle. In the two following sections 4 and 5 we will show oracle inequalities for the new method.

## 2. Prerequisites

For Tikhonov regularization it always holds that  $\|x - x_{n+1}\| \leq \|x - x_n\|$ . We will need a slightly more powerful inequality at this point and therefore use a set of assumptions introduced in [BK08].

**Assumption 2.1** *Let  $x$  such that either (see [BK08] eq. (9), slightly rewritten)*

$$\|(A^*A)^{-1}x\|_{\mathcal{X}} < \infty$$

*or that there exist constants  $\gamma > 0$ ,  $1 > \nu > 0$ ,  $C_{\gamma,\nu} > 0$  and  $D_{\gamma,\nu} > 0$  such that for all  $0 \leq t \leq \gamma$  (see [BK08] Definition 2.2)*

$$D_{\gamma,\nu}^2 t^{2\nu} \geq \sum_{\{k:|\sigma_k^2 \leq t\}} \langle x, u_k \rangle^2 \geq C_{\gamma,\nu}^2 t^{2\nu}$$

**Remark 2.2** *There are other functional analytic formulations of this assumption (cf. [KN08]); however, the general idea is similar; namely that  $x$  should have a rather uniform distribution of the energy in its coefficients  $\langle x, u_k \rangle$ .*

**Lemma 2.3** (cf. [BK08]) *Let  $x$  fulfill assumption 2.1. Then it holds*

$$\|x - x_{n+1}\| \leq w_1 \|x - x_n\| \tag{6}$$

*with  $0 < w_1 < 1$ .*

Now we will introduce two different noise models. Classically, one considers deterministic noise

**Definition 2.4 (Deterministic Noise)**  $\xi$  is called deterministic noise of noise level  $\delta$  if

$$\|\xi\| \leq \delta$$

*The noise behavior  $\rho(\cdot)$  wrt the regularization parameter  $n$  is defined as*

$$\rho(n) = \|A_n^{-1}\| \delta \leq \delta (q_0 q^n)^{-1/2}$$

As all the results can be transferred easily to the case of colored noise by modifying the function  $\rho$  we will now just consider white noise.

**Definition 2.5 (Stochastic Noise)** Let  $\xi$  a white noise Gaussian random variable, i.e., the  $\langle x, u_k \rangle$  are independent and identically distributed (iid) along the distribution  $\mathcal{N}(0, \delta^2)$ . The noise behavior  $\rho(\cdot)$  wrt the regularization parameter  $n$  is defined as

$$\rho(n)^2 = \mathbb{E}\|A_n^{-1}\xi\|^2 = \delta^2 \text{trace}(A_n^{-1})$$

In both cases this trivially yields

**Lemma 2.6** There exists  $1 < w_2$  such that

$$\rho(n) \leq \rho(n+1) \leq w_2 \rho(n) \quad (7)$$

**Remark 2.7** Almost every other constant in this article will be based on  $w_1$  and  $w_2$ . Though desirable we cannot give a general upper or lower bound for these constants, they are purely problem dependent. At first glance this might look as a disadvantage, however this means as well that we get problem specific optimal results.

### 3. Fast Balancing

The question of the optimality of a regularization parameter  $n_{opt}$  is rather difficult in a deterministic setting. The most natural definition, namely

$$n_o = \underset{n}{\operatorname{argmin}} \|x_n^\delta - x\|,$$

would be best. However, there is no known concept which leads to proofs in a general setting. Thus we will use the second-best solution. Using the triangle inequality, the error  $\|x_n^\delta - x\|$  is bounded by the sum of a decreasing function  $\|x_n - x\|$  and an increasing function  $\|A_n^{-1}\xi\|$  which itself is bounded by  $\rho(n)$ .

$$\|x_n^\delta - x\| \leq \|x_n - x\| + \|A_n^{-1}\xi\| \leq \|x_n - x\| + \rho(n)$$

Again, the point

$$n_{oo} = \underset{n}{\operatorname{argmin}} \|x_n - x\| + \rho(n)$$

is inaccessible. However, using the definition (8) below we can at least guarantee that

$$2(\|x_{n_{oo}} - x\| + \rho(n_{oo})) \geq \|x_{n_{opt}} - x\| + \rho(n_{opt}).$$

Interestingly, the stochastic case is much easier due to the independence of  $\xi$  and  $x$ :

$$\mathbb{E}\|x_n^\delta - x\|^2 = \|x_n - x\|^2 + \rho(n)^2 + 2 \left\langle \left( A_n^{-1} \right)^* (x_n - x), \xi \right\rangle = \|x_n - x\|^2 + \rho(n)^2.$$

In this case, the parameter  $n_{opt}$  as defined below is really optimal on average:

**Definition 3.1 (Optimal Parameter)** The optimal parameter  $n_{opt}$  is defined such that

$$\|x_{n_{opt}} - x\| > \rho(n_{opt}) \quad \text{and} \quad \|x_{n_{opt}+1} - x\| \leq \rho(n_{opt} + 1) \quad (8)$$

**Remark 3.2** This parameter, of course, just exists when the noise level  $\delta$  is sufficiently small. However, if this is not the case the noise described by  $\rho(\cdot)$  dominates the information  $x$  so much that 0 would be the best regularized solution. It is common practice to assume that this parameter  $n_{opt}$  exists.

Now we define the Fast and the Lepskij balancing principle and perform a first comparison

**Definition 3.3 (Balancing Functional)** Let  $k \geq 1$ . The balancing functional is defined as

$$b_k(n) = \max_{n < m \leq n+k} \left\{ 4^{-1} \|x_n^\delta - x_m^\delta\| \rho(m)^{-1} \right\} \quad (9)$$

**Definition 3.4 (Lepskij Balancing Principle)** Let  $b_k(n)$  defined as in (9). Furthermore let  $\tau \geq 1$  and  $N \geq n_*$ . The Lepskij Balancing Parameter  $n_L(\tau, N) = n_L$  is defined as

$$n_L(\tau, N) = \operatorname{argmin}_n \{b_k(m) < \tau \quad \forall N \geq m \geq n\} \quad (10)$$

**Definition 3.5 (Fast Balancing Principle)** Let  $b_k(n)$  defined as in (9) and let  $\tau \geq 0$ . The fast balancing parameter

$n_* = n_*(\tau)$  is defined as

$$n_*(\tau) = \operatorname{argmin}_n \{b_k(n) < \tau\} \quad (11)$$

**Remark 3.6** In contrast to the Lepskij balancing principle [MP03, MP06, BM07], no upper bound  $N$  is needed.

**Lemma 3.7** It holds for all admissible pairs  $(N, \tau)$

$$n_L(\tau, N) \geq \min\{N, n_*(\tau)\}$$

### Proof

This is a direct consequence out of (10) and (11).  $\square$

**Remark 3.8** For the sake of simpler notation we will mostly refer to  $n_L(\tau, N)$  by  $n_L$  and to  $n_*(\tau)$  by  $n_*$ . Just when the particular choice of  $N$  and  $\tau$  is importance we will keep these parameters.

The condition that  $N \geq n_*$  is of course rigorously seen not fulfillable without knowing the optimal regularization parameter, however this is a standard assumption for all proofs for the Lepskij balancing principle and does not pose any particular problems in practice where  $N$  is normally chosen anyway in the range of the machine precision.

## 4. Deterministic Case

Now we will show that – as for the Lepskij balancing principle – for the fast balancing principle an oracle inequality holds.

**Theorem 4.1** Let  $n_*$  and  $n_{opt}$  defined as above. Furthermore assume assumption 2.1 to be valid. It holds for the constants  $w_1, w_2$  defined in (6) and (7) independent of  $\delta$

$$\|x_{n_*}^\delta - x\| \leq C \left( \|x_{n_{opt}} - x\| + \rho(n_{opt}) \right)$$

where  $C = \frac{(4\tau+3)}{2} \min_{1 \leq \kappa \leq k} \left\{ \frac{w_2^\kappa}{1-w_1^\kappa} \right\}$ .

### Proof

Assume  $m > n > n_{opt}$ . Due to assumption 2.1 and either definition of the noise the equations (6) and (7) hold and we obtain with the triangle inequality

$$\|x_n^\delta - x_m^\delta\| \leq \|x_n - x\| + \|x_m - x\| + \rho(n) + \rho(m) < 4\rho(m)$$

and hence

$$b_k(n) < 1$$

and thus  $n_* \leq n_{opt}$  and hence  $\rho(n_*) \leq \rho(n_{opt})$  due to (7). Using (8) we have

$$\|x_{n_*} - x\| \geq \rho(n_*).$$

On the other hand, for  $n \leq n_*$  and  $\kappa \leq k$ , using the inverse triangle inequality and (6) resp. (7):

$$\begin{aligned} 4\tau\rho(n+\kappa) &\geq \|x_n^\delta - x_{n+\kappa}^\delta\| \\ &\geq \|x_n - x\| - \|x_{n+\kappa} - x\| - \rho(n) - \rho(n+\kappa) \\ &\geq (1-w_1^\kappa)\|x_n - x\| - 2\rho(n+\kappa). \end{aligned}$$

Hence

$$\|x_{n_*} - x\| \leq (4\tau+2) \min_{1 \leq \kappa \leq k} \left\{ \frac{w_2^\kappa}{1-w_1^\kappa} \right\} \rho(n_*)$$

and so using (7)

$$\begin{aligned} \|x_{n_*}^\delta - x\| &\leq \|x_{n_*} - x\| + \rho(n_*) \\ &\leq (4\tau+3) \min_{1 \leq \kappa \leq k} \left\{ \frac{w_2^\kappa}{1-w_1^\kappa} \right\} \rho(n_*) \\ &\leq (4\tau+3) \min_{1 \leq \kappa \leq k} \left\{ \frac{w_2^\kappa}{1-w_1^\kappa} \right\} \rho(n_{opt}) \\ &\leq \frac{(4\tau+3)}{2} \min_{1 \leq \kappa \leq k} \left\{ \frac{w_2^\kappa}{1-w_1^\kappa} \right\} (\|x_{n_{opt}} - x\| + \rho(n_{opt})). \end{aligned} \quad \square$$

**Remark 4.2** Obviously, we obtain the best result for the minimal admissible  $\tau$ , i.e.  $\tau = 1$ . Furthermore,  $C$  is independent of  $\delta$  as  $w_1$  and  $w_2$  are.

## 5. Stochastic Case

We will need a stochastic bound for the probability that the observed error is much bigger than our estimation  $\rho(\cdot)$ .

**Lemma 5.1** (see e.g., [BR08]) *Let  $Z = \sum_{k=1}^{\infty} \alpha_k^2 \zeta_k^2$  with  $\sum_{k=1}^{\infty} \alpha_k^2 = 1$  and  $\zeta_k \sim N(0, 1)$  iid. Assume that  $\max_k \alpha_k > 0$ . Then*

$$\forall z > 0 : \mathbb{P}(Z \geq z) \leq \sqrt{2}e^{-z/4}. \quad (12)$$

The proofs would become far too complicated if we used  $b_k(\cdot)$  for any  $k$ . Therefore, we will restrict our attention to the case  $k = 1$ , i.e.,  $b_1(\cdot)$ . In numerical implementations we observed that the results improve for slightly bigger  $k$ . From a certain  $k$  onwards (problem-dependent) it turns out that  $k$  does not seem to have an influence on the solution any more.

**Lemma 5.2** *Assume assumption 2.1 to be valid. For  $n < n_{opt}$  there exist constants  $c_1, c_2 > 0$  depending on  $w_1$  and  $w_2$  but not depending on  $\delta$  and  $\tau$  such that*

$$\mathbb{P}\{b_1(n) < \tau\} \leq c_1 \exp(-c_2(n_{opt} - n)^2) \exp(\tau^2) \quad (13)$$

*Furthermore for  $n > n_{opt}$  there exists a constant  $c_3 > 0$  depending on  $w_1$  and  $w_2$  but not depending on  $\delta$  and not depending on  $\tau$  such that*

$$\mathbb{P}\{b_1(n) \geq \tau\} \leq c_3 \exp(-\tau^2/4). \quad (14)$$

### Proof

Let  $n < n_{opt}$ . Then using (3), (4), (5), (6), (7), (8), (12) and the triangle inequality we

get (usage of the equations marked on top)

$$\begin{aligned}
& \mathbb{P} \left\{ 4^{-1} \|x_n^\delta - x_{n+1}^\delta\| \rho(n+1)^{-1} < \tau \right\} \\
&= \mathbb{P} \left\{ 4\tau\rho(n+1) > \|x_n^\delta - x_{n+1}^\delta\| \right\} \\
&\stackrel{ti}{\leq} \mathbb{P} \left\{ 4\tau\rho(n+1) > \|x_n - x\| - \|x_{n+1} - x\| - \|x_n^\delta - x_n\| - \|x_{n+1}^\delta - x_{n+1}\| \right\} \\
&\stackrel{(6)}{\leq} \mathbb{P} \left\{ 4\tau > \frac{(1-w_1)\|x_n - x\|}{\rho(n+1)} - \frac{\|x_n^\delta - x_n\|}{\rho(n+1)} - \frac{\|x_{n+1}^\delta - x_{n+1}\|}{\rho(n+1)} \right\} \\
&= \mathbb{P} \left\{ \frac{\|x_n^\delta - x_n\|}{\rho(n+1)} + \frac{\|x_{n+1}^\delta - x_{n+1}\|}{\rho(n+1)} > \frac{(1-w_1)\|x_n - x\|}{\rho(n+1)} - 4\tau \right\} \\
&\stackrel{(6)}{\leq} \mathbb{P} \left\{ \frac{\|x_n^\delta - x_n\|}{\rho(n+1)} + \frac{\|x_{n+1}^\delta - x_{n+1}\|}{\rho(n+1)} > \frac{(1-w_1)w_1^{-(n_{opt}-n)}\|x_{n_{opt}} - x\|}{\rho(n_{opt})} - 4\tau \right\} \\
&\stackrel{(3)(4)(5)(7)(8)}{\leq} \mathbb{P} \left\{ \frac{\|A_n^{-1}\xi\|}{\rho(n)} + \frac{\|A_{n+1}^{-1}\xi\|}{\rho(n+1)} > (1-w_1)w_1^{-(n_{opt}-n)} - 4\tau \right\} \\
&\leq \mathbb{P} \left\{ \frac{\|A_n^{-1}\xi\|}{\rho(n)} > \frac{1-w_1}{2}w_1^{-(n_{opt}-n)} - 2\tau \right\} + \mathbb{P} \left\{ \frac{\|A_{n+1}^{-1}\xi\|}{\rho(n+1)} > \frac{1-w_1}{2}w_1^{-(n_{opt}-n)} - 2\tau \right\} \\
&\stackrel{(12)}{\leq} 2\sqrt{2} \exp \left( - \left( \frac{1-w_1}{2}w_1^{-(n_{opt}-n)} - 2\tau \right)^2 / 4 \right) \\
&\leq 2\sqrt{2} \exp \left( - \left( \frac{1-w_1}{4}w_1^{-(n_{opt}-n)} \right)^2 \right) \exp(\tau^2) \\
&\leq c_1 \exp(-c_2(n_{opt}-n)^2) \exp(\tau^2)
\end{aligned}$$

for some appropriate constants  $c_1$  and  $c_2$ . Now let  $n > n_{opt}$ . Then using (3), (4), (5), (7), (8), (12) and the triangle inequality we get

$$\begin{aligned}
& \mathbb{P} \left\{ 4^{-1} \|x_n^\delta - x_{n+1}^\delta\| \rho(n+1)^{-1} \geq \tau \right\} \\
&= \mathbb{P} \left\{ 4\tau\rho(n+1) \leq \|x_n^\delta - x_{n+1}^\delta\| \right\} \\
&\stackrel{ti}{\leq} \mathbb{P} \left\{ 4\tau\rho(n+1) \leq \|x_n - x\| + \|x_{n+1} - x\| + \|x_n^\delta - x_n\| + \|x_{n+1}^\delta - x_{n+1}\| \right\} \\
&\stackrel{(7)(8)}{\leq} \mathbb{P} \left\{ (4\tau - 2)\rho(n+1) \leq \|x_n^\delta - x_n\| + \|x_{n+1}^\delta - x_{n+1}\| \right\} \\
&\stackrel{(3)(4)(5)(7)}{\leq} \mathbb{P} \left\{ 2\tau \leq \frac{\|A_n^{-1}\xi\|}{\rho(n)} + \frac{\|A_{n+1}^{-1}\xi\|}{\rho(n+1)} \right\} \\
&\leq \mathbb{P} \left\{ \tau \leq \frac{\|A_n^{-1}\xi\|}{\rho(n)} \right\} + \mathbb{P} \left\{ \tau \leq \frac{\|A_{n+1}^{-1}\xi\|}{\rho(n+1)} \right\} \\
&\stackrel{(12)}{\leq} 2\sqrt{2} \exp(-\tau^2/4). \quad \square
\end{aligned}$$

**Lemma 5.3** Assume assumption 2.1 to be valid and constants  $c_1, c_2$  as defined in lemma 5.2. Let  $k = 1$ . Then for  $n < n_{opt}$

$$\mathbb{P} \{n_* = n\} \leq c_1 \exp(-c_2(n_{opt}-n)^2) \exp(\tau^2). \quad (15)$$

**Proof**

It holds:

$$\begin{aligned}\mathbb{P}\{n_* = n\} &= \mathbb{P}\{b_1(n) < \tau \text{ and } \forall_{m < n} b_1(m) > \tau\} \\ &\leq \mathbb{P}\{b_1(n) < \tau\} \leq c_1 \exp(-c_2(n_{opt} - n)^2) \exp(\tau^2).\end{aligned}\quad \square$$

The situation for  $n > n_{opt}$  becomes much more complicated. It all depends on the question how fast the probabilities  $\mathbb{P}\{b_1(n) \geq \tau\}$  and  $\mathbb{P}\{b_1(m) \geq \tau\}$  decorrelate. Therefore, we will first present both the extreme cases and then discuss their implications.

**Lemma 5.4** *Assume assumption 2.1 to be valid and  $c_3$  defined as in lemma 5.2. Let  $n > n_{opt}$ . Then*

$$\mathbb{P}\{n_* \in \{n_{opt} + 1 \dots n\}\} \leq c_3 \exp(-\tau^2/4). \quad (16)$$

*Assume additionally that for all  $m < n$  it holds  $\mathbb{P}\{b_1(n) \geq \tau\} \text{ and } b_1(m) \geq \tau\} = \mathbb{P}\{b_1(n) \geq \tau\} \mathbb{P}\{b_1(m) \geq \tau\}$ . Then*

$$\mathbb{P}\{n_* = n\} \leq (c_3 \exp(-\tau^2/4))^{n-n_{opt}}. \quad (17)$$

**Proof**

Direct consequence of lemma 5.2.  $\square$

In the case of decorrelation, the likelihood of the event  $n_* > n_{opt}$  decreases very fast, whereas in the worst case (i.e. the perfectly correlated case) the likelihood is constant. Please note that in any case we are in a considerably better situation than with the Lepskij balancing principle, where, in dependence of an upper bound  $N$ , the probability is  $N \exp(-\tau^2/4)$ .

In reality we cannot expect complete decorrelation of events, however considering practical observations the following relaxed condition seems to be reasonable:

**Assumption 5.5** *There exists a constant  $C_3(w_1, w_2) > c_3$  such that for  $n > n_*$*

$$\mathbb{P}\{n_* = n\} \leq (C_3 \exp(-\tau^2/4))^{n-n_{opt}}. \quad (18)$$

*Assume furthermore that  $\tau$  big enough such that*

$$C_3 \exp(-\tau^2/4) w_2 < 1 \quad (19)$$

**Remark 5.6** *Obviously we have no justified way to find out which  $\tau$  fulfills (19); however as  $C_3$  just depends on  $w_1$  and  $w_2$  and not on  $\delta$  we also have this independence from  $\delta$  for  $\tau$ .*

*In most practical situations  $\tau = 1$  seems to be sufficient and a dependence of  $\delta$  has not been observed (which might of course just be due to the limited range of numbers processable on modern computer hardware).*

**Theorem 5.7** Assume that assumptions 2.1 and 5.5 are valid. Then we have

$$\mathbb{E}\|x_{n_*}^\delta - x\|^2 \leq c_5 \left( 1 + c_6 \exp(\tau^2/2) + c_5 \frac{1}{1 - C_3 \exp(-\tau^2/8) w_2} \right) (\|x_{n_{opt}} - x\|^2 + \rho(n_{opt})^2)$$

with variables  $c_5$  and  $c_6$  just depending on  $w_1$  and  $w_2$ .

### Proof

Due to the independence of  $x$  and  $\xi$  and an inequality connecting the different moments of Gaussian random variables (see e.g., [BR08]), we have for  $c_5 = (4\Gamma(3))^{1/4}$

$$\begin{aligned} \mathbb{E}\|x - x_n^\delta\|^4 &= \|x - x_n\|^4 + 2\|x - x_n\|^2 \mathbb{E}\|x_n - x_n^\delta\|^2 + \mathbb{E}\|x_n - x_n^\delta\|^4 \\ &\leq \|x - x_n\|^4 + 2\|x - x_n\|^2 \mathbb{E}\|x_n - x_n^\delta\|^2 + (c_5 \mathbb{E}\|x_n - x_n^\delta\|^2)^2 \\ &\leq (\|x - x_n\|^2 + c_5 \mathbb{E}\|x_n - x_n^\delta\|^2)^2 \\ &\leq c_5^2 (\|x - x_n\|^2 + \mathbb{E}\|x_n - x_n^\delta\|^2)^2 \\ &= c_5^2 (\mathbb{E}\|x - x_n^\delta\|^2)^2. \end{aligned}$$

For an appropriate constant  $c_6$  independent of  $\tau$  it holds using (6), (7), (8), (13), (18) and the Hölder inequality:

$$\begin{aligned} \mathbb{E}\|x_{n_*}^\delta - x\|^2 &= \sum_{n=1}^{\infty} \mathbb{E} (\|x - x_n^\delta\|^2 \mathbf{1}_{n=n_*}) \\ &\leq \sum_{n=1}^{n_{opt}-1} (\mathbb{E}\|x - x_n^\delta\|^4)^{1/2} (\mathbb{E}\mathbf{1}_{n=n_*}^2)^{1/2} \\ &\quad + \mathbb{E}\|x_{n_{opt}}^\delta - x\|^2 \\ &\quad + \sum_{n=n_{opt}+1}^{\infty} (\mathbb{E}\|x - x_n^\delta\|^4)^{1/2} (\mathbb{E}\mathbf{1}_{n=n_*}^2)^{1/2} \\ &\stackrel{(13)(18)}{\leq} \sum_{n=1}^{n_{opt}-1} c_5 (\|x - x_n\|^2 + \rho(n)^2) (c_1 \exp(-c_2(n_{opt} - n)^2/2) \exp(\tau^2/2)) \\ &\quad + \|x - x_{n_{opt}}\|^2 + \rho(n_{opt})^2 \\ &\quad + \sum_{n=n_{opt}+1}^{\infty} c_5 (\|x - x_n\|^2 + \rho(n)^2) (C_3 \exp(-\tau^2/8))^{n-n_{opt}} \\ &\stackrel{(6)(7)(8)}{\leq} \sum_{n=1}^{n_{opt}-1} c_5 (w_1^{-(n_{opt}-n)} \|x - x_{n_{opt}}\|^2 + \rho(n_{opt})^2) (c_1 \exp(-c_2(n_{opt} - n)^2/2) \exp(\tau^2/2)) \\ &\quad + \|x - x_{n_{opt}}\|^2 + \rho(n_{opt})^2 \\ &\quad + \sum_{n=n_{opt}+1}^{\infty} c_5 (\|x - x_{n_{opt}}\|^2 + w_2^{n-n_{opt}} \rho(n_{opt})^2) (C_3 \exp(-\tau^2/8))^{n-n_{opt}} \\ &\leq \|x - x_{n_{opt}}\|^2 c_5 \left( 1 + c_6 \exp(\tau^2/2) + \frac{1}{1 - C_3 \exp(-\tau^2/8)} \right) \\ &\quad + \rho(n_{opt})^2 c_5 \left( 1 + c_6 \exp(\tau^2/2) + c_5 \frac{1}{1 - C_3 \exp(-\tau^2/8) w_2} \right), \end{aligned}$$

which proves the assertion.  $\square$

Even if we cannot assume decorrelation (assumption 5.5), we still have a convergence result due to  $n_L(\tau) \geq \min\{N, n_*(\tau)\}$ , also compare [RH08].

**Theorem 5.8** *Assume that assumptions 2.1 to be valid. Let  $n_* = \min\{N, n_*(\tau)\}$  and  $\tau = c \log(\delta^{-1})$ . Then it holds*

$$\mathbb{E}\|x_{n_L}^\delta - x\|^2 \leq C_L \left( \|x_{n_{opt}} - x\|^2 + \log(\delta^{-1})\rho(n_{opt})^2 \right)$$

and

$$\mathbb{E}\|x_{n_*}^\delta - x\|^2 \leq C_* \left( \|x_{n_{opt}} - x\|^2 + \log(\delta^{-1})\rho(n_{opt})^2 \right).$$

### Proof

The first part was proven in [BP05]; the second part is trivial, using that on the one hand  $n_L(\tau) \geq \min\{N, n_*(\tau)\}$ . On the other hand, the risk for  $n_*$  being too small (which is not affected by the decorrelation effect) can be bounded from above by a multiple of  $\|x_{n_{opt}} - x\|^2 + \rho(n_{opt})^2$  as shown in the last theorem.  $\square$

## 6. Conclusion

In numerical experiments [BL10] it is almost impossible to distinguish the results of the Lepskij balancing principle and the newly introduced fast balancing. This can be interpreted as follows:

- There is just a very low probability for outliers in the noise which influence the noise behavior  $\rho(\cdot)$  after the optimal regularization parameter.
- There is an extremely low probability that the noise modifies the data in such a way that one stops too early.

However, considering the computation time, the new method has big advantages, it can compete easily with other methods like the Morozov discrepancy principle.

In conclusion, we have shown that this modification of the Lepskij balancing principle is very well suited for practice and should replace it in all time-critical applications as long as one does not need to fear big frequency gaps in the solutions.

This analysis has been done only for Tikhonov regularization. A similar analysis imposing stricter requirements on the solution  $x$  was performed for truncated Singular Value Decomposition (TSVD) in [Bau10].

Furthermore, large numerical experiments [BL10] show that the newly defined method works very well and, in contrast to most other parameter choice regimes, can cope with colored noise without any performance loss. In these experiments it was observed that the factor  $C$  in the oracle inequality is at most around 2. The method is very stable, i.e., the number of observed outliers is very low, both for Tikhonov and Spectral-Cut-Off regularization. This behavior does not change when one replaces the exact noise behavior by an estimation based on several measurements [Bau10].

## Acknowledgements

The author gratefully acknowledges the financial support by the Upper Austrian Technology and Research Promotion.

## References

- [Bau10] F. Bauer, *Applying lepskij-balancing in practice*, Arxiv.org (2010).
- [BH05] F. Bauer and T. Hohage, *A Lepskij-type stopping rule for regularized Newton methods*, Inverse Problems **21** (2005), no. 6, 1975–1991.
- [BHM09] F. Bauer, T. Hohage, and A. Munk, *Iteratively regularized Gauss–Newton method for nonlinear inverse problems with random noise*, SIAM Journal on Numerical Analysis **47** (2009), no. 3, 1827–1846.
- [BK08] F. Bauer and S. Kindermann, *The quasi-optimality criterion for classical inverse problems*, Inverse Problems **24** (2008), 035002.
- [BL10] F. Bauer and M. Lukas, *Comparing parameter choice methods for regularization of ill-posed problems*, Mathematics and Computers in Simulation, Preprint: <http://www.bmath.de/Docs/mainLowRes.pdf> (submitted 2010).
- [BM07] F. Bauer and A. Munk, *Optimal regularization for ill-posed problems in metric spaces*, J. Inverse Ill-Posed Probl. **15** (2007), no. 2, 137–148.
- [BP05] F. Bauer and S. Pereverzev, *Regularization without preliminary knowledge of smoothness and error behavior*, European Journal of Applied Mathematics **16** (2005), no. 3, 303–317.
- [BR08] F. Bauer and M. Reiß, *Regularization independent of the noise level: an analysis of quasi-optimality*, Inverse Problems **24** (2008), no. 5, 055009 (16pp).
- [EHN96] H. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Kluwer Academic Publisher, Dordrecht, Boston, London, 1996.
- [GP00] A. Goldenshluger and S. Pereverzev, *Adaptive estimation of linear functionals in hilbert scales from indirect white noise observations*, Probab. Theory Related Fields **118** (2000), 169–186.
- [Hof86] B. Hofmann, *Regularization of applied inverse and ill-posed problems*, Teubner, Leipzig, 1986.
- [HPR07] U. Hämarik, R. Palm, and T. Raus, *Use of extrapolation in regularization methods*, J. Inverse Ill-Posed Probl. **15** (2007), no. 3, 277–294.
- [HPR08] ———, *Extrapolation of tikhonov and lavrentiev regularization methods*, Journal of Physics **135** (2008), 012048 (8pp).
- [KN08] S. Kindermann and A. Neubauer, *On the convergence of the quasi-optimality criterion for (iterated) tikhonov regularization*, Inv. Prob. Imag. **2** (2008), no. 2, 291–299.
- [Lep90] O.V. Lepski, *On a problem of adaptive estimation in gaussian white noise*, Theory of Probability and its Applications **35** (1990), no. 3, 454–466.
- [MP03] P. Mathé and S. Pereverzev, *Geometry of linear ill-posed problems in variable hilbert spaces*, Inverse Problems **19** (2003), no. 3, 789–803.
- [MP06] Peter Mathé and Sergei V. Pereverzev, *Regularization of some linear ill-posed problems with discretized random noisy data.*, Math. Comput. **75** (2006), no. 256, 1913–1929 (English).
- [RH08] T. Raus and Hämarik, *About the balancing principle for choice of the regularization parameter*, Journal of Physics **135** (2008), 012087 (8pp).